

## How Do we Make Electronic Archives Usable and Accessible?

Paper prepared for Documenting the Digital Age, San Francisco, February 10 - 12, 1997

Margaret Hedstrom, School of Information, University of Michigan

Let me begin my remarks with a description of the access process in many electronic archives today.

A junior in college is working on a research paper about Gulf War Syndrome among women. She has read newspaper and journal articles, two books on the topic, and several government reports. She would like to find some quantitative data so that she can compare women and men. She also read that there was an on-line discussion group of women who served in the Gulf War and she would like to find its archive to analyze how they have reacted to Gulf War Syndrome.

She finds several on-line catalogs from repositories of electronic records and identifies potentially useful sources. She sends an e-mail request to one repository where the staff photocopy the finding aid and user documentation for each data set of interest and mail it to her. After reviewing the documentation, she selects two files of interest and faxes an order for them. One file is available on a floppy disk; the other is available only on magnetic tape. The archives has a three day backlog of copy requests. When her request reaches the front of the queue, archives staff copy the requested files and mail them to the student. Three weeks elapse between the student's initial interest and receipt of the data. The student has to locate a computing facility on campus that maintains a tape drive. She has to reformat the data and arrange to transfer it through the campus network to her personal workstation. She cannot find the address or any information about the e-mail discussion group so she decides to abandon that part of her analysis.

Meanwhile, the archives staff has compiled data on the use of its collection. The use figures are appalling. Although they receive several hundred e-mail questions each month asking about specific data sets, most requesters lose interest when they learn that they have to purchase data on diskettes or magnetic tape and wait for it to be copied and shipped. Administrators are asking: why are we keeping all of this stuff when no one uses it?

Although this vignette does not describe the only way that electronic archives are made accessible today, it illustrates all too common problems with access to archival materials in electronic form. Locating potentially useful sources can be frustrating and undependable because access tools are not comprehensive or integrated into a uniform access system. Retrieving materials from off-line storage and delivering them in antiquated formats is time

consuming and labor intensive for both the repository and the researcher. Little used archives are difficult to justify, especially when ongoing investments in physical maintenance of the collection are necessary to avoid physical deterioration and technological obsolescence.

More importantly, this vignette illustrates the relationships between accessibility, convenience, levels of use, and the costs of delivering services. While there are no comprehensive analyses of the use patterns among established electronic archives, anecdotal evidence suggests that the more readily accessible the materials, the more likely they are to be used.<sup>1</sup> Moreover, making electronic records accessible on-line can be more cost-effective for the repository than producing custom made copies on demand. We should use several criteria to devise and select strategies that will make electronic archives accessible and usable. First, we should identify approaches to access that best satisfy users' needs. Second, we should consider how to provide access to electronic archives at a reasonable cost and in a more economic manner than is common in archives today. Third, we should make certain that providing access to electronic archives will improve the process and results of research. Accessibility is imperative for electronic archives, not only to meet rising user demands and expectations, but to develop an economically sustainable model of archival services.

My conception of "electronic archives" is a highly distributed one. In thinking about accessibility and usability, I envision a world where a wide variety of institutions and individuals will take responsibility for preserving digital information. Some electronic archives will be maintained by special repositories dedicated exclusively to preserving and providing access to digital information; others will be extensions of traditional archives with hybrid collections of paper-based and electronic records. Many valuable electronic sources, however, will be made available directly by their original creator or producer because it is impractical to transfer custody to a special repository or because the institution or individual who created the records has ongoing need for them.<sup>2</sup> We need to develop strategies and methods for accessibility and usability that can span a variety of custodial arrangements.

The use of computer and network technologies to disseminate descriptive information about archival records and to provide remote access to their contents shows promise of vastly improving access to archival records. National and international databases, such as the Research Libraries Information Network (RLIN) maintained by Research Libraries Group (RLG), contain catalog records that describe more than half a million archival collections in repositories around the world. The archival community is in the final stages of developing and promulgating a standard for Encoded Archival Description which uses SGML to produce browsable and searchable on-line finding aids for special collections.<sup>3</sup> These are important building blocks in the development of comprehensive and integrated access systems for archival materials, although much remains to be done to realize their full potential. Only a small percentage of all archival records are described in network-accessible databases, and most descriptions only provide access at the very general level

of the collection or records series. Only a minuscule portion of current archival records have machine-readable finding aids, indexes, and other access tools that help researchers locate specific documents or items, and only a tiny portion of current holdings have been converted to digital formats for network delivery.

One of the first steps that we can take to make electronic archives accessible is to integrate descriptive information about them into existing access systems for archives, special collections, and other primary source materials. This is important for several reasons. Users should be able to locate electronic sources through local, national, and international access systems without having to search separate catalogs or databases which are segregated by format or form of material. As electronic archives become comprised of multifarious and highly heterogeneous types of information, segregation by format (electronic versus non-electronic) will present obstacles to accessibility. It made sense to establish special access systems for machine-readable data archives when the term "machine-readable" was largely synonymous with numeric and statistical data files. Now, electronic archives can contain any form of material -- textual documents, photographs, sound, moving images, maps, drawings, or data. We still live in a hybrid environment where many processes are only documented adequately through a combination of electronic and paper sources. Maintaining linkages between different formats of materials will become increasingly burdensome if we do not find ways to develop integrated access systems. Multi-media products defy categorization by format, and I would urge us to avoid the temptation to establish yet another type of archive --the multi-media archive.

We should take the notion of integrating electronic archives into access systems for traditional archival materials one step further by investigating ways to integrate or link access systems for archives (paper and electronic) with the access systems for cultural resources that reside in traditional and digital libraries, museums, and other cultural institutions. Access systems for electronic archives should allow users to navigate through layers of increasingly detailed description that will help them identify, locate, and evaluate primary source material. Making electronic resources available on-line will not obviate the need for cataloging and descriptive information about each resource, and it may, in fact, make such information even more critical. Access systems should provide core descriptive elements that identify the origin or creator of the source, its title, inclusive dates, extent, and some minimal level of subject access. The Dublin Core metadata elements, developed through a workshop sponsored by OCLC and the National Center for Supercomputing Applications, offers one model for such a high-level directory that could support discovery of network resources.<sup>4</sup> This level of description should lead to a more detailed finding aid, appropriate for archival materials, which describes the scope and content of the resource, provides evaluation criteria, and explains the origin and provenance of the source. For some resources, detailed indexes linked to the finding aid would provide access to files, documents, or specific items. For some types of material, technical documentation should be provided detailing such attributes as the file

structure, coding or representation schemes, hardware and software requirements, or other features of the source. Users could navigate through these layers of description to identify and select materials relevant to their problem or research question.

Before we go much farther with designing access systems, we need a systematic and comprehensive analysis of users and their requirements. Any effort to make electronic archives accessible and usable will be hindered by the lack of knowledge about current and potential users of archives. Even without the introduction of on-line access systems and networked resources, the user community for archival materials has become increasingly diverse in recent decades. Once the sanctum of historians and other scholars, archives have become known by and appealing to a larger, more popular, and more diverse user population. Alex Haley's book *Roots* is accredited with fueling a nascent movement of avocational researchers seeking records for genealogy and family history.<sup>5</sup> The use of compelling primary source materials as illustrations in books or as the basis for documentary films, such as Ken Burns' documentaries on the Civil War and Baseball, introduce primary source materials to large and popular audiences. Archival materials have played an increasingly central role in uncovering evidence from the past that supports legal claims against violations of civil rights or implied contracts, reveals patterns of negligence, or establishes linkages between exposure to certain agents and medical consequences which can have life threatening effects.<sup>6</sup> Teachers have begun to work with archivists to select archival materials for use in the curriculum because students find primary sources engaging and they provide excellent tools for learning how to evaluate and interpret various forms of evidence.

These observations speak in part to the broad perspective that we must maintain when considering what to keep in electronic archives, but I am raising these points to address the question of accessibility. Although the general trends I described above are reinforced by anecdotal evidence of changing user needs and by scattered statistics from reading rooms, the archival community does not have a good understanding of its current or potential user community, their interests, their facility for using and understanding primary source material, or their needs. When we add to this the potential for making electronic archives accessible to a much larger user community, with different needs and abilities, sometimes without human mediation we add another layer of complexity to the question of accessibility. I would argue strongly for systematic studies of why users seek archival materials, what mechanisms they use to discover sources, how satisfied they are with the materials they find, how much they are willing to invest in finding and gaining access to archival materials, which delivery mechanisms they prefer, and what problems they encounter in using and interpreting the sources they find. Such research would be useful only if it were extended beyond the current user population to identify potential and future users whose needs may differ considerably from those of the current user population. Without such data, we will not be able to design access systems that address user needs effectively.

Building electronic archives that are accessible to a wide variety of users in the formats they most prefer is only half the battle. We will have accomplished little if we cannot deliver sources that are usable by requesters. There are numerous options and tools for delivering electronic documents to users with Internet access, but many of these approaches are not robust enough to deliver reliable, authentic, and usable archival records. The characteristics of archival records as documentary evidence of human activity demand specific strategies and management methods that will protect their integrity while enhancing access to their contents.<sup>7</sup>

One of the primary concerns is that most archival records have to be presented in a larger context because they rarely can stand alone as unique, bounded objects that are self explanatory. Contextual information about the creator, purpose, events surrounding the creation of a record, and its chain of custody is essential for determining the reliability of electronic documents and for interpreting their contents. The principle of provenance remains at the core of strategies for managing archives in the network environment. Respect for the principle of provenance means that archival records must not be separated conceptually from the broader context of their origin, creation, and use. Contextual information, which is critical for interpreting the contents of archival records in any format, includes knowledge of the relationships among documents, the circumstances that gave rise to their creation, their intent or purpose, their receipt and use, and the chain of custody from the originator to the present custodian.

Contextual information can be provided through a variety of means. Specific metadata that explicitly describes the context from which archival sources were derived can be attached or linked to each record or document. Structural elements can be imbedded in the documents to provide visual and other semiotic clues about their origins. Adhering to the principle of provenance often demands examination of legal mandates or bureaucratic regulations which require the creation of certain types of records, biographical research about individuals, and knowledge of the administrative history, organizational structure, and business processes of the entities that generate records. At least at the outset, the people who are building electronic archives will have to make a concerted effort to capture or supply sufficient contextual information about the contents of digital archives because much of the digital information being generated today is not self-documenting. Our culture is inventing new forms of documentary evidence that are technologically complex, yet socially and culturally primitive. Electronic mail systems, for example, deliver a variety of messages in an undifferentiated structure, ranging from formal transactions, to deeply personal communications between friends, to anonymous postings on bulletin boards. Document conventions have not evolved sufficiently to support effective management of electronic records or consistent interpretation of their contents. Documentary forms are becoming more sophisticated and refined, however, with increasing possibilities for creating self-referential documents, and archivists are beginning to understand the core descriptive elements that must accompany content to make it meaningful.

Recent research on electronic records management has identified metadata models and elements that should accompany digital objects to support their authentication and long-term management. Although there is no single model or set of metadata specifications, several initiatives have proposed ways to attach metadata to electronic documents or files in order to address problems of authentication, interpretation, and archiving. One such model, developed in a research project at the University of Pittsburgh, divides descriptive information about electronic records into six categories:

- 1) registration metadata which uniquely identified each electronic object;
- 2) terms and conditions metadata that contains information about access restrictions or other condition of use;
- 3) structural metadata with information about the file or document structure;
- 4) contextual metadata with information about the creation and provenance of the document;
- 5) content metadata describing the logical and physical aspects of the content; and
- 6) metadata on use of each record.<sup>8</sup>

Presently, archivists would have to extract, compile, and structure this metadata because few systems have been designed to supply and organize metadata in a consistent standardized manner. If models such as this become widely adopted, however, one can envision a time when more electronic records will be self-documenting.

We should also develop the means to distribute the software needed to open, view, and analyze electronic materials with the records themselves. The problem of software dependency and software obsolescence is one of the most intractable obstacles facing electronic archives. Few archives have the technical resources to maintain obsolete versions of software that might be required to open, view, and manipulate archival records which were created using software that has been updated or replaced. The notion of a distributed electronic archive offers a partial solution to this problem. It should be technically feasible for a few sites to maintain older versions of software or emulators of older versions that run on the current generation of hardware and operating systems. Users needing access to older software in order to use electronic records in obsolete formats would be able to download and install the software on their own workstations or submit requests to a server that supports the software. Such an approach would serve a dual purpose. It would provide users with access to software tools that are difficult to locate and install and it would provide a means to preserve software as an important intellectual and cultural resource in its own right. This approach will not eliminate the need for periodic migration of electronic records because eventually the incompatibilities between older software and current hardware and operating systems will become insurmountable. Nevertheless, this strategy could reduce the frequency of migrations, provide access to records with the same look and feel as their original format, and curb maintenance and migration costs.

There is a great deal that archivists and designers can do to build electronic archives that are accessible and useable, but we should be cautious about placing all of the functionality into the archival system itself. Adequate descriptive information and techniques like time/date stamps and encryption, can be employed to prevent alteration of records. But we will need to launch a parallel effort to teach the users of electronic archives how to be discriminating and skeptical consumers of digital information. Learning how to evaluate and interpret evidence has always been an implicit goal of our educational system. While the specific skills needed to evaluate digital documents may differ from those used for older forms of records, they are no less essential. Here we can learn from the experience of European scholars and archivists who, upon discovering that many medieval documents were fakes and forgeries, developed the discipline of diplomatics in the seventeenth century to analyze and authenticate documents.<sup>9</sup> Some archivists today are applying the principles of diplomatics to digital information with the intent of building into modern information systems the capability of producing reliable and authentic records. But we must also think about ways to teach users the principles of a new digital diplomatics so that they can apply these principles themselves to make educated judgments about the accuracy, reliability, and authenticity of the documents that they retrieve from electronic archives. We need to educate the next generation of scholars as well as the general public how to approach digital evidence with a questioning mind about how it was generated, why it was preserved, and how it might be interpreted. Until we feel as comfortable with electronic evidence as we do with traditional forms of documentation, archivists will have a responsibility to help users evaluate, understand, and interpret new documentary forms.

The actions taken by individuals and organizations to save and care for their own archives will play a vital role in enriching the archival record. We should pursue strategies that change the norms of individual record keeping, allow people to build their own digital archives, increase awareness of the practical and cultural value of documentary evidence, and developing simple tools that help individuals and organizations save and protect their records. Current software tools that "save" or "archive" documents, whether designed for individuals using microcomputers or for complex networks, fall short of what is needed to capture and preserve meaning-rich records. While personal and organizational collections of digital materials might be turned over to specialized archives at some point, the ability of archival repositories to provide meaningful access to such collections will depend to a large extent on the measures that the original creators take to organize, describe and care for their records. To the extent possible, recordkeeping standards and practices should be integrated into the processes of records creation and maintenance, support the access and retrieval requirements of the records creator, and protect the integrity and authenticity of records.<sup>10</sup>

No discussion of accessibility and usability would be complete without raising the issues of affordability and access restrictions. Increasing concerns about personal privacy, efforts to gain or retain control over intellectual

property, and the growth of fee-based access services all work against widely accessible electronic archives. Archivists and researchers will not be able to shape individual or societal norms about privacy and access to personal or confidential information. Yet there are some practical measures that the developers of digital archives can take to mitigate privacy concerns and support legitimate access to private or confidential information. Any electronic archives should develop comprehensive policies that define the terms and conditions for release of records, the degree of access restrictions acceptable to the archives, and the requirements for use of restricted sources. Prior to acquiring or gaining control over materials, the archives should negotiate with each donor a clear statement of access restrictions. Some archives will need to develop redaction capabilities that mask individual identities or permit the selected release of portions of files or documents. In developing policies for access, electronic archives can learn much from the experience of repositories of traditional formats of materials. Respectable archives have formal access policies and the archival profession as a whole embraces the principle that restrictions on access should be kept to a minimum. If access restrictions are necessary to comply with privacy or other access restrictions or to secure donations of materials, access restrictions should apply equitably to all users to specific categories of users.<sup>11</sup>

The law and policies around intellectual property and user fees will also be decided by forces outside the archival community. Nevertheless, developers of electronic archives must be cognizant of the impact of intellectual property issues on both the usability of the archive and the complexity of its administration. From the user's perspective, electronic archives should encourage, if not require, donors to place their materials in the public domain. If this is not possible, the archives should negotiate for liberal fair use provisions. Regardless of the outcome of such negotiations, it will be essential for the archives to carefully document the copyright status of its holdings and the provisions for requesting permission to use materials that are subject to copyright. Likewise, electronic archives should resist the temptation to impose user fees for personal, scholarly, or educational uses of the archives. In building electronic archives, we are creating a cultural resource and serving a larger public good. While charges for the commercial use of the archives might provide one source of revenue, we should not subordinate the larger social and cultural objectives of electronic archives to their commercial viability.

Let me close with an alternative vision of using electronic archives:

The college junior with her research topic on women and the Gulf War Syndrome in mind searches a high-level directory using natural language to describe her research question and define the types of sources of interest. The search returns a list of eighteen possible sources at five different sites ranked by relevance to her selection criteria. She is most interested in the third and fourth items on the list and asks for additional information about them. The search returns the full text of the finding aid and a database with all of the data elements in each file. She

searches these to discover that only one of the data files breaks down the data by gender. She then looks at other attributes and discovers that this source is a complete registry of Gulf War veterans who have been treated for Gulf War Syndrome. She can download a public use version of the file which includes data on each case but does not include personal identifiers. She requests the file and four minutes later, it resides on her hard drive. The initial search also listed the address of an e-mail discussion group of women afflicted with Gulf War Syndrome with instructions about how to access the archive. She had not considered a source like this, but now decides to use it as well to analyze how women are responding to Gulf War Syndrome.

The archive is keeping detailed statistics on requests and use of its collection. They notice is that those sources which can be downloaded directly by users are almost fifty times more likely to be used than those that have to be ordered and shipped using off-line media. They also use the statistics on requests to decide which types of sources to pursue. They have noticed a fifteen-fold increase in requests since they started the remote access service, but since most of these requests are self-serve the demand for technical services has actually declined. The reference staff is very busy answering e-mail and helping users interpret their data. The head of the archives uses these statistics, along with several letters from requesters praising the service, to make the case to his Board that this is a valuable service. He secures an increase in funding that will be used to hire more reference staff and put more collections on-line.

This is the future that we should strive for in electronic archives. In order to achieve this vision, we will need to enhance and link access systems so that electronic records are widely known or easily discoverable through the access systems that requesters normally use when seeking archival materials. We will have to develop the means to deliver materials as seamlessly as possible, with the minimum restrictions on reuse, and at little or no cost. The objects that are delivered will be useful only if they are accompanied by or can be linked to rich resources of descriptive and contextual information. This contextual information will help end users assess the quality, reliability, and relevance of the documents to their problem or question. Pointers will help them find similar or related materials if they wish to delve further into the electronic archive or find relevant print sources. But we should not expect to build all of the selection and evaluation capabilities into the archive itself. We must also educate users to become discriminating consumers of archival materials and critical readers of electronic evidence.

## Notes

<sup>1</sup>Thomas J. Ruller, "Open All Night: Using the Internet to Improve Access to Archives. A Case Study of the New York State Archives and Records Administration," forthcoming in *Reference Services for Archives and Manuscripts*, Laura B. Cohen, ed., New York: Haworth Press, 1997.

<sup>2</sup> The issue of physical custody of electronic records is the subject of an ongoing debate among archivists. For recent discussions of this question see David Bearman, "An Indefensible Bastion: Archives As a Repository in the Electronic Age," in *Archival Management of Electronic Records*, David Bearman, ed., Archives and Museum Informatics Technical Report, No. 13 (1991), 14-24; Kenneth Thibodeau, "To Be or Not To Be: Archives for Electronic Records," in *Archival Management of Electronic Records*, 1-13; Margaret Hedstrom, "Archives: To Be or Not To Be? A Commentary," in *Archival Management of Electronic Records*, 25-30; Terry Cook, "Leaving Archival Electronic Records in Institutions: Policy and Monitoring Arrangements at the National Archives of Canada," *Archives and Museum Informatics* 9:2 (1995): 141-49; and the November 1996 issue, *Archives and Manuscripts* (vol. 24, no. 2), the journal of the Australian Society of Archivists which was devoted to the issue of custody and post-custodial archives.

<sup>3</sup> The authoritative web site for information about the EAD is <http://lcweb2.loc.gov/ammem/ead/>.

<sup>4</sup> For information about the Dublin Core and the subsequent Warwick Framework, see [www.oclc.org:5046/conferences/metadata/dublin\\_core\\_report.html](http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html) and Lorcan Dempsey and Stuart Weibel, "The Warwick Metadata Workshop: A Framework for the Deployment of Resource Description," *D-Lib Magazine* (July/August 1996): [www.dlib.org/dlib/july96/07weibel.html](http://www.dlib.org/dlib/july96/07weibel.html).

<sup>5</sup> Alex Haley, *Roots*, Garden City, N.J.: Doubleday, 1976.

<sup>6</sup> One example of this type of resource is the Comprehensive Epidemiologic Data Resource developed by the U.S. Department of Energy to provide public access to data about health and exposure data at DOE installations. See <http://cedr.lbl.gov/>.

<sup>7</sup> Margaret Hedstrom, "Electronic Archives: Integrity and Access in the Network Environment," in *Networking in the Humanities: Proceedings of the Second Conference held at Elvetham Hall*, Hampshire, UK, 13-16 April 1994, Kent: Bowker-Saur, 1995: 77-95.

<sup>8</sup> David Bearman and Ken Sochats, "Metadata Requirements for Evidence" (1995): [www.lis.pitt.edu/~nhprc.html](http://www.lis.pitt.edu/~nhprc.html).

<sup>9</sup> Don C. Skemer, "Diplomatics and Archives," *American Archivist* 52 (Summer 1989): 376-82; and Luciana Duranti, "Diplomatics: New Uses for an Old Science," Part VI, *Archivaria* 33 (Winter 1991-92): 6-24.

<sup>10</sup> A good example of effective recordkeeping standards and practices is the Statement of Common Position on Electronic Recordkeeping entitled "Corporate Memory in the Electronic Age." The statement, issued in May 1996, was produced by a meeting of key industry participants, individual practitioners, and professional organizations, sponsored by the Australian Council of Archives in Sydney on October 23, 1995 [www.aa.gov.au/AA\\_WWW/ProAssn/ACA/Corpmenw.htm](http://www.aa.gov.au/AA_WWW/ProAssn/ACA/Corpmenw.htm)

<sup>11</sup> Society of American Archivists/American Library Association, "Joint Statement on Access to Original Research Materials," *American Archivist* 42:4 (Fall 1979).